Digital Connection

Don Rotolo, N2IRZ

The Global Positioning System

From CQ Magazine, January 2016

Many of us use GPS almost every day, and we all know what it does for us: providing our geographic location on the Earth. As a byproduct, it also tells us the time. Some use GPS to find their way through the web of roads around us, while others just want to know exactly where they are.

Curiosity may have killed the cat, but when you stop learning, you die. In that vein, this month we'll examine how this GPS thing works. Perhaps you'll never need to parse or modify the GPS message protocol, or nudge a satellite's orbit, but as I've said many times before, a little knowledge won't hurt you.

Where the heck are we?

Let's start with the basics of location. If I am on a flat surface and can measure an accurate distance from one fixed point (whose location is known), I can be certain that I am somewhere on a circle surrounding that known point. If I can then measure from a second point, then I can be certain my location is one of the two points where those two circles intersect – quite a bit more precise than just a circle.

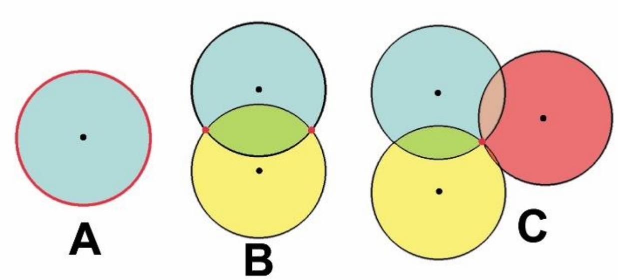

Figure 1: On two dimensions, if we know the distance from a single point our possible locations are a circle (A). Two distances (B) narrow it down to 2 points where they intersect. The third distance allows us to know which of the two points is our location.

Ah, before we go further, I need to make something clear: Accuracy and precision are not the same thing. Accuracy is how close something is to the true value, while Precision is smallest increment that is measured. As an example a weatherman in Fairbanks, Alaska says "On New Year's Day, the high temperature will be between -60°F and 75°F". He is very accurate – historically the temperature should be around zero – but he is not at all precise. If he were instead to say "On New Year's Day, the high temperature will be 53.802°F", he is very precise (down to a thousandth of a degree!) but probably not at all accurate. I'll be using these terms quite a lot in this month's column, and it is important you understand the difference.

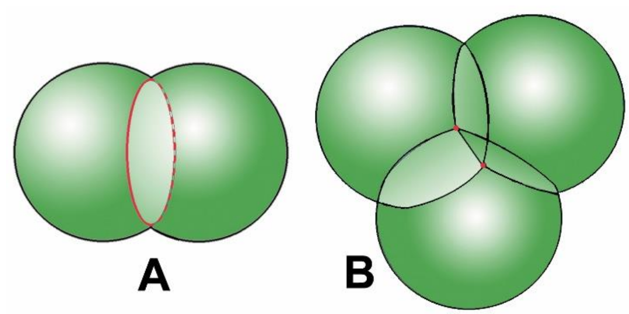

Figure 2: Adding a third dimension increases complexity slightly: Two spherical distances (A) tell us we are located somewhere on a circle, while three distances (B) narrows our location down to two points. If we were to add a fourth distance we would know which of the two points represented our exact location.

OK, back to locations: So we have our location narrowed down to one of two points. How accurately we know our location depends on how accurately we can measure the distances, and of course how accurately we know the known locations. If we wanted to determine which of those two points we are at, we can just take a third measurement from some third known point, which unambiguously provides our exact location.

Three distance measurements provides a single point on a flat plane, but what happens if we expand our world to three dimensions? In such a case, two measurements provide a circle

of possible locations – same as our single measurement on the flat plane – and three measurements provide two possible points – same as our two measurements before. Considering that, it becomes clear that all we need is a fourth measurement to determine which of the two points we are at, and we can then know our location in three dimensions.

In a sense, that's what GPS does: It measures the distance from four (or more) known points to calculate a location in three dimensions. But the implementation is somewhat more complicated, because we can't directly measure the distances from a satellite (my longest tape measure is 100 feet), and those satellites are moving.

One nice thing about orbital mechanics is that you can calculate and predict the location of an object in space like a satellite both accurately and precisely. Johannes Kepler figured out the laws of planetary motion (technically, celestial mechanics) around 1600 CE, which Isaac Newton refined and generalized almost a century later. Modern orbital mechanics are based on Albert Einstein's General Theory of Relativity, which is more exact and universally applicable than Newton's Laws.

Space operations

The United States has a global network of ground facilities and people who track the GPS satellites and make sure we all know where they are (or will be), allowing for accuracies down to a few tens of feet for most users. With certain augmentations, it is possible to get down into the millimeter range. In fact, GPS is used to track the movement of earth's tectonic plates, which move up to about 4 inches a year.

Since we can't directly measure distance to a satellite we instead use time to calculate distances. Each satellite has a very accurate Rubidium (atomic) clock on board, which is kept accurate by periodic corrections from the ground. (Because of the effects of time slowing in the presence of a gravitational field described by General Relativity, the clocks on the GPS satellites run about 38 microseconds a day faster than those on earth). The satellites send their version of the time in each data message. Easy enough.

Along with the time, each satellite sends its orbital path data (known as ephemeris), along with the satellite ID number and some housekeeping information. Using the ephemeris and the current time, the receiver calculates the satellite's position and can then determine how far away the satellite is. Knowing how far away they are, the receiver calculates backwards to determine the propagation delay from each satellite. The delay values are used to adjust the receiver's internal clock, leaving us with an easy way to calculate the distance to each satellite – the propagation speed of radio waves. With four or more distances from known locations, we can know where we are.

This can also be done with only 3 satellites, but only if the exact time is known. Since atomic clocks are expensive, adding a fourth satellite allows us to use a far less accurate receiver clock. The receiver's clock only needs to hold its accuracy for a short time, as new data is sent every six seconds, so it is relatively inexpensive while able to be both accurate and extremely precise – a few parts per billion (PPB).

On a side note, TAPR's Totally Accurate Clock < http://www.tapr.org/kits_tac2.html> used this feature of GPS receivers, along with a crystal oscillator, to deliver a time signal with around 1 PPB accuracy, previously almost unattainable by a regular citizen. Unfortunately, the TAC kit is no longer available, but can be reproduced with some effort, using the information found on the TAPR website.

Once these calculations are started, the location of the receiver is determined. The actual accuracy depends on many factors, including variables like ionospheric propagation delay, so short-term one can expect a location within maybe 50 feet or so. From personal experience, the calculated location drifts around over time, so some days my iPad shows the blue dot right on my house, and sometimes it's in the backyard, or over the neighbor's fence. Many years ago, I experimented with a raw GPS receiver and found that over several days the indicated position formed a circle of sorts, with the most likely actual position being more or less at the center of the circle.

The standard GPS signal, known as L1 C/A (Coarse Acquisition), is what almost everyone uses. There are other signals (described below), but not all satellites transmit them at this time. In the early days of GPS, the government intentionally distorted the signal with a system known as Selective Availability (SA). It was believed that inaccuracies in the location data would help foil any evil uses of GPS. It was later decided to eliminate SA, I believe because the powers that be realized that it didn't make much difference (at least as far as Evil goes) and civilians have been getting full single-band accuracy since then.

More recently, in addition to C/A, two new civilian signals were added: L1C and L2C. L1C is on the L1 band (1575 MHz), where the C stands for Civilian, and L2C is on the L2 band (1227

MHz). Having the second signal greatly enhances accuracy, particularly by allowing for ionospheric propagation distortion detection and correction. For the military, the data is encrypted and somewhat more precise. For the future, testing is underway for a new L5 (1176 MHz) signal to be used for Safety-of-Life signals, the idea being for precision aircraft operations and the like.

If you're familiar with Amateur Radio Digital modes, then this will make sense to you: GPS satellites transmit Spread Spectrum signals using BPSK (Binary Phase-Shift Keying) modulation, with a Code-Division Multiple-Access (CDMA) scheme to allow the reception of several satellites by a single receiver. Each signal is modulated by a different Pseudo-Random Number (PRN), which means in practice that it is fairly easy to distinguish the signal of one satellite from another, even though they are being received through the same antenna. Forward-Error Correction (FEC) is added to increase signal robustness.

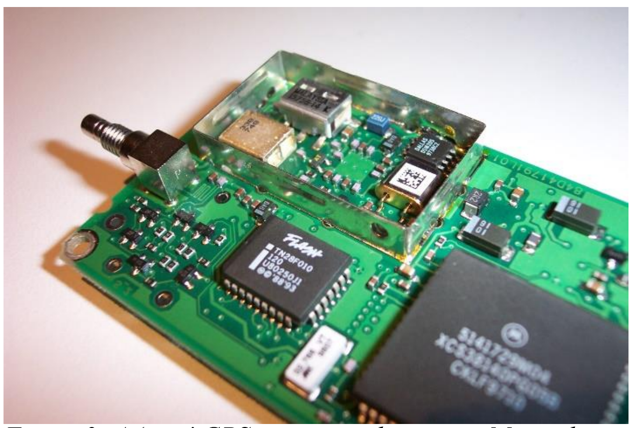

Figure 3: A 'raw' GPS engine, in this case a Motorola product used in the automotive industry. This example has the oscillator 'can' cover removed so the components inside are visible. At left is the antenna connector, not seen at the other end are the connections for power, ground and received data.

The receiver also picks the best quality data and weights that more heavily as it does its calculations. The term for 'goodness' of data is called DOP, or Dilution Of Precision. Four satellites bunched together directly overhead will have a low level of accuracy or a high DOP, while four satellites widely spaced apart in the sky would deliver a low DOP. The ideal case would be one satellite directly overhead and three others spaced widely apart somewhat above the horizon. There are different values for Horizontal DOP, Vertical DOP and a combination of the two called Positional DOP. A DOP of 1 or less is about as good as it gets, while a value of 10 is almost unusable.

Different GPS signals have their own format, but they have mostly the same components. The L1 C/A signal that most of us use transmits five 300-bit data subframes, each consisting of 10 words. The first subframe contains the satellite's time signal, and the second and third contain the satellite's ephemeris. The last two subframes contain a small chunk of the Almanac, which is a complete description of the entire GPS satellite constellation – coarse orbital data, satellite status messages, and an ionospheric model used for propagation time corrections. The Almanac is not as precise as the ephemeris, but the receiver needs this information to determine which satellites should be visible to it so it can start looking for them. It takes around 12.5 minutes to receive a complete almanac, but it remains valid and useable for quite some time (weeks), while ephemeris data needs to be hears fairly often (every few hours at least).

Practical GPS

Most people's interaction with GPS involves navigating their automobile. GPS receiver manufacturers add software to their product to not only perform the navigation functionality, but tweaks the data so the vehicle icon is displayed atop the nearest road, a so-called Augmentation of the data.

Users who cannot accept an error of several feet – like a land surveyor – can get access to higher precision with other augmentations. For example, if you set a receiver into a known place on earth, you can compare the actual position with what GPS is telling you and measure the amount of error. Using a local radio transmitter, broadcast that error signal over a few miles, and a surveyor can get accuracies of a fraction of an inch. Another method is to average the position data over time which, like my experiment many years ago, tends to cancel out the errors. There are other ways – including a way to use some of the encrypted military data (legally!) - but I'll leave finding out how that's done up to you.

NMEA

The GPS satellite signal is academically interesting, but down here on Earth, we can play around with the data spewing from our GPS receiver. This data stream is standardized in the NMEA (National Marine Electronics Association) format, but some GPS Engines (a raw, unpackaged GPC receiver module) have their own proprietary format for use in specialized devices.

NMEA is a 4800 Baud 8N1 serial data stream that is designed to drive an EIA-422 device, but can also drive an RS-232 interface. The GPS

receiver sends one NMEA string, or sentence, every second. Sentences are predefined message strings that contain specific data, a common one being the GGA sentence. Always starting with $GP, a typical data sentence might look like:

$GPGGA,082519,3403.268,N,8418.389,W,1,08,0.9,346.2,M,96.4,M,,*27

Although the specification is widely published for you to find, the four key data points are the time (08:25:19 UTC), the latitude (34°° 3.236' N) and longitude (84° 18.389W), and the altitude (346.2 meters). The "1" just after the W is an indicator of the quality of the positional fix data (The 1 is typical as a 'good GPS fix'), the 08 is the number of satellites being tracked, and the 0.9 is the Horizontal Dilution of Precision (HDOP).

Those so inclined could buy a GPS engine on the Internet for maybe ten dollars, although I saw one for under a dollar brand new with antenna. Parse the NMEA string with your Arduino or microcontroller of choice and consider the possibilities.

This month we dipped a little into the theoretical, since last time we got fairly practical with our discussion of FSQ. Next time, I'm not exactly sure where we'll go: I am hoping to line up a very special interview, but these things are never sure so I won't promise anything or even mention the details. Of course, if YOU have something you think I should cover, be sure to write, or even if you just want to tell me where to go. With GPS, I suppose...

Until next time,

73 de N2IRZ

Sidebar: That time of year again

Happy New Year 2016. It's the time of year where tens of thousands of high school students (and their mentors) learn the new FIRST Robotics Competition (FRC) game. In only SIX weeks, students around the world must design, build and program a 150 pound robot to play whatever this year's game will be. If you have some time or talent that you can share, find a local FRC team at <http://www.usfirst./org> and volunteer some time to help our next generation of engineers and scientists learn real-world skills...and maybe win a competition against other local teams. After 25 years, and over 3000 teams worldwide, they must be doing something right.

###